Published online June 20, 2006

Evaluation of *Glycine* Germplasm for Nulls of the Immunodominant Allergen P34/Gly m Bd 30k

Leina M. Joseph, Theodore Hymowitz,* Monica A. Schmidt, and Eliot M. Herman

ABSTRACT

Soybean [*Glycine max* (L.) Merr.] seed contains an immunodominant human allergen P34 or Gly m Bd 30k (mentioned as P34) of the cysteine protease family. Of approximately 16 266 accessions from USDA soybean germplasm screened, 12 P34 null lines were identified among soybean (*G. max*), wild annual (*Glycine soja* Sieb. and Zucc.), and wild perennial *Glycine* spp. *Glycine soja* were low P34 expressers, while *G. max* and wild perennial species had nondetectable levels of the allergenic protein. Further investigation of *G. max* nulls by 2D-IEF/SDS PAGE showed all primary seed proteins present indicating that the loss of P34 was not due to large scale restructuring of protein content. Southern and northern analysis showed no large insertions or deletions to render the gene nonfunctional. The cDNA of both *G. max* nulls each showed the same six point mutations indicating the two nulls have a single origin. Of these six single nucleotide changes, four are predicted to result in an amino acid alteration. One such alteration results in a serine being replaced by a cysteine residue. The introduction of a cysteine residue might produce a mismatched disulfide bond formation producing an unstable P34 protein in the null soybean accessions. The isolation and introgression of soybean lines with low allergen levels will provide the basis for developing a low allergen line incorporated with other agronomically desirable traits in a breeding program.

SOYBEAN is grown primarily for production of seed, which has a multitude of uses in the food and industrial sectors. It represents one of the major sources of edible vegetable oil and protein in the food industry. However, soybeans contain allergenic proteins, making it one of the major food sources that cause allergies to sensitive individuals (Herman, 2004). In the USA and Europe, nearly 5 to 8% of babies and 2% of adults are reported to be allergic to soybeans (Heppel et al., 1987); therefore, millions of infants (babies) must avoid soybean-based formula and baby food (Cantani and Lucenti, 1997). Further, the increased use of soy proteins and their derivatives in a wide range of processed foods limits soy-sensitive individuals in their choices of soy free processed foods. Currently, the primary treatment for food allergies is to avoid the food source causing the allergy. Although recently enacted labeling laws will assist in identifying foods with soy content, it will still be difficult to avoid soy protein because of its extensive usage in a wide range of processed foods (Herian et al., 1993; Tsuji et al., 1995). Soybean allergic reaction in humans is primarily atopic (skin) reactions and gastric distress which can range from mild to severe and is rarely life-threatening, although there are some reports of soy-induced anaphylaxis (Foucard and Malmheded, 1999).

Soybean possesses about 15 proteins recognized by IgEs from soy-sensitive people (Burks et al., 1989; Ogawa et al., 1991, 1993). The dominant soybean allergens consist of P34 and the α-subunit of conglycinin. P34 is a unique member of the papain superfamily of cysteine proteases lacking the catalytic cysteine residue that is replaced by a glycine (Kalinski et al., 1992; Herman, 2005). It is most strongly and frequently recognized by the IgE antibodies in sera of soy-sensitive patients with atopic dermatitis (Ogawa et al., 1991). In several IgE binding studies, more than 65% of soy-sensitive patients react only to P34 protein (Ogawa et al., 1991, 1993; Helm et al., 1998, 2000), suggesting that the 34-kDa protein is the major allergen in soybean and a target allergen for producing low-allergen content hypoallergenic soybean. Epitope mapping of P34 showed at least 12 distinct epitopes on P34 protein through immunological analysis of its allergenicity (Helm et al., 1998 as reported by Yaklich et al., 1999). P34 is a relatively minor seed constituent comprising less than 1% of total seed protein. The lack of P34 in soybean would eliminate a large portion of the allergenicity of soybean seeds.

The function of P34 in soybeans is not known with the exception of its role in syringolide elicitor binding activity in disease resistance (Cheng et al., 1998). The P34 protein consists of 379 amino acids having homology to thiol proteases (Kalinski et al., 1990). In soybean seed, P34 is post translationally processed from a 46-kDa precursor protein after seed germination. Seed thiol proteases are synthesized only after seed germination, whereas P34 accumulates during seed maturation. Electron microscope immunochemistry with a monoclonal antibody demonstrated that P34 is localized in the protein storage vacuoles but not in the oil bodies (Kalinski et al., 1992). It is moderately abundant in soybean seed and cotyledons but is found in low levels in leaves.

Using cosuppression gene silencing technique, a transgenic soybean was produced in which P34 was eliminated (Herman et al., 2003). These soybean plants had apparently identical seed composition, development, structure, and phenotypic characters as that of nontransgenic plants, indicating that P34 protein is not required for seed protein processing or maturation. However, regulatory difficulties and the lack of acceptance of genetically modified soybeans by the baby food and formula industry makes using such an allergen-suppressed soybean difficult at the present time.

An alternative approach would be to identify low-allergen content soybeans in the soybean germplasm. An initial attempt to find soybeans lacking P34 by screen-

L.M. Joseph and T. Hymowitz, Dep. of Crop Sciences, Univ. of Illinois, Turner Hall, 1102 S. Goodwin Ave., Urbana, IL 61801; M.A. Schmidt and E.M. Herman, USDA-ARS, Donald Danforth Plant Science Center, 975 N. Warson Rd., St. Louis, MO 63132. Drs. Joseph and Schmidt contributed equally to the work. Received 27 Feb. 2006. *Corresponding author (soyui@uiuc.edu).

Published in Crop Sci. 46:1755–1763 (2006).
Plant Genetic Resources
doi:10.2135/cropsci2005.12-0500

677 S. Segoe Rd., Madison, WI 53711 USA

ing a limited number of domesticated and wild soybean relatives was negative (Yaklich et al., 1999). A similar survey of Japanese lines did not produce any P34 null lines (Ogawa et al., 2000). In this paper, we have extended this approach by screening the entire USDA national soybean germplasm collection (except Japanese lines that has been previously surveyed) with the aim of identifying a naturally occurring soybean with reduced or, preferably, absent P34. We report here that we have identified two soybean and three wild perennials that are P34 nulls and seven *G. soja* in which some are nulls and some have significantly low levels of P34. The two *G. max* accessions lacking the P34 protein are characterized.

MATERIALS AND METHODS

Seed Material

Approximately 16 266 accessions from the USDA *Glycine* collection (Urbana, IL) consisting of soybean (*G. max*), its wild annual (*G. soja*), and wild *Glycine* perennial relatives were screened to identify accessions with low or no P34 expression. Individual seed of each accession was used as starting material to extract total protein as the allergen of interest is expressed highly in seeds. Seeds of the wild annual and perennial relatives of soybean were weighed to 0.5 g for the assay because of their smaller size.

Seed Processing and Immunoblotting for P34 Assay

Pulverized seed from each accession was added to 1 mL TBS-T buffer [0.05 *M* Tris, 0.5 *M* NaCl, pH 7.4 with 0.05% (v/v) Tween 20 (polyoxyethylene sorbitan monolaurate)] and incubated for 30 min on a shaker (200 rpm) at room temperature.

Immuno-blot assay of the seed extracts was done by a rapid and large scale method as follows: A 10-μL aliquot of each sample extract was dot-blotted on a nitrocellulose membrane (0.45 μm; Fisher Scientific) with a P34 positive control sample (cv. Kunitz). To check for equal loading of the samples on the blots, membranes were stained with Ponceau S solution (Sigma, St. Louis, MO) for 1 min and rinsed in double distilled water. The membrane was incubated in TBS buffer (0.05 *M* Tris and 0.5 *M* NaCl, pH 7.4) for 5 min and blocked (10% nonfat dry milk powder in TBS buffer) for 1 h followed by rinsing twice in TBS buffer for 5 min each. The membranes of *G. max* and *G. soja* seed extracts were incubated in P34 protein specific monoclonal antibody P4B5 (diluted 1: 5000 in TBS-T buffer with 0.1% milk powder) for 1 h. The blots of wild perennial *Glycine* spp. were incubated in P34 specific monoclonal antibody P3E1 (Herman et al., 1990). The membranes were washed extensively in TBS-T buffer and incubated in the secondary antibody [antimouse IgG (Fab specific) alkaline phosphatase conjugate (Sigma)] diluted 1: 10 000 in TBS-T buffer with 0.1% milk powder for 1 h and rinsed in TBS-T buffer and TBS buffer once each. The P34 protein was detected by a color substrate system [BCIP/NBT: final concentrations 0.02% (w/v) 5-bromo-4-chloro-3-indolyl phosphate and 0.03% (w/v) nitro blue tetrazolium in 70% (v/v) dimethylformamide] (Sigma).

One-Dimensional Protein Analysis

Twelve P34 null accessions identified from the screen were analyzed by one-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE; Laemmli, 1970). Total soluble protein was extracted from soybean seeds according to Sambrook et al. (1989). Protein content was determined by Bradford (1976) assay. Five and 10 μg of protein was loaded onto 4 to 18% (w/v) SDS-PAGE gradient gels and electrophoretically separated for a western blot and gel staining, respectively. Seed extract from cv. Kunitz was included as a positive control. Protein gels were stained in 0.1% (w/v) Coomassie Brilliant Blue R250 in 40% (v/v) methanol, 10% (v/v) acetic acid overnight, and destained for approximately 3 h in 40% methanol and 10% acetic acid. Replicate gels for all three samples were electroblotted to Immobilon-P transfer membranes (Millipore, Bedford, MA) for Immunoassay analysis. Blots were blocked (3% milk in TBS) overnight at 4°C and incubated with 1:1000 dilution monoclonal antibody P4B5 specific to the P34 protein (Herman et al., 1990) for 1 h at room temperature. Blots were washed in TBS buffer and incubated with the secondary phosphate-conjugated antimouse antibody (1:10 000 dilution) for 1 h. Antibody detection was performed using chromogenic substrate as described above. The same immunoassay procedure was done on a replicate gel with a second monoclonal antibody P3E1 that binds to a different active site from that of first monoclonal antibody P4B5. Two different monoclonals were used for the test to ensure that nonbinding of monoclonals to protein was not due to the absence of a single active site in case of a altered P34 protein.

Two-Dimensional Protein Analysis

Total protein was isolated from mature seeds according to a modified phenol method (Hurkman and Tanaka, 1986; Hajduch et al., 2005). Seeds (0.5 g) from three soybean accessions (Kunitz, *G. max* PI603570A, and *G. max* PI567476) were ground in liquid nitrogen and resuspended in 0.5 mL of Tris buffered phenol (pH 8.8) and 0.5 mL of extraction buffer containing 0.1 *M* Tris-HCl pH 8.8, 10 m*M* EDTA, 0.4% (v/v) 2-mercaptoethanol, and 0.9 *M* sucrose. Protein pellets were washed as previously described (Hajduch et al., 2005) and solubilized in a DeStreak rehydration solution (GE Healthcare, Pittsburgh, PA). Protein was quantified by Bradford assay. Total protein extract (150 μg) was loaded onto an 11 cm immobilized pH gradient (IPG) gel strips (pH 3–10 nonlinear) (BioRad, Hercules, CA) and allowed to hydrate overnight. Isoelectric focusing (IEF) was performed for a total of 40 kVh with a Protean IEF Cell (BioRad). The IPG strips were equilibrated according to manufacturer's procedures (BioRad). Second dimension SDS-PAGE gels (8–16% linear gradient) were run on a Criterion Cell (BioRad) for 15 min at 60 V and then at 200 V for 1 h. Blotting and subsequent immunodetection with a P34-specific monoclonal antibody were as previously described.

Southern Analysis

Genomic DNA was isolated from leaf tissue according to the CTAB (cetyl trimethyl ammonium bromide) method of Doyle and Doyle (1990). Ten micrograms of genomic DNA was digested with either *Hae*III or *Eco*RI, separated on a 0.8% (w/v) agarose gel for approximately 11 h and subsequently transferred to a positively charged nylon membrane (Amersham, Buckinghamshire, UK) using 0.4 *M* NaOH as both a transfer and denaturing agent (Sambrook et al., 1989). Blots were hybridized in a modified Church's buffer (Church and Gilbert, 1984) consisting of 7% (w/v) SDS, 0.5 *M* sodium phosphate pH 7.2, and 0.5 m*M* ethylenediaminetetracetic acid (EDTA). A P34 gene specific probe was produced with Rediprime II random priming labeling kit (Amersham) with [α-^{32}P] dCTP (New England Nuclear) and 25 ng of a 1.1-kb amplicon of P34 as a template. Unincorporated nucleotides were removed with a Micro Bio spin 6 chromatography column (BioRad) and the probe denatured at 95°C for 5 min and hybridized overnight

at 65°C. Membranes were exposed to BioMax autoradiography film (Kodak, Rochester, NY) for 2 d at −80°C.

The hybridization probe was produced by PCR in 25-μL final volume that contained 200 μ*M* each of dATP, dCTP, dTTP and dGTP and 1m*M* $MgCl_2$, 1× buffer (ABI), 1 μL of 1 U μL^{-1} VENT polymerase (New England Biolabs, Ipswich, MA), 2 μL of cDNA, 5 μ*M* of the upstream P34 primer corresponding to the region from base 1 to 23 (sequence 5′-TT ATGGGTTTCCTTCTGTTGCTT-3′); and 5 μ*M* of the downstream P34 primer corresponding to bases 1148 to 1175 (sequence 5′-CAAGCACTCGTTGTATTGAACCTTTAA-3′). PCR cycling parameters were 94°C for 10 min and then 45 cycles of 94°C for 30 s, 55°C for 45 s, and 72°C for 1 min, followed by a final extension at 72°C for 10 min. The PCR product was isolated by the gel extraction kit (Qiagen, Valencia, CA). (Bases are numbered according to GenBank accession 05560).

Northern Analysis

Total RNA was isolated as described by Cashmore et al. (1978) with the following modifications. Briefly, five 150-mg cotyledons from the two P34 null accessions of interest (*G. max* PI603570A and *G. max* PI567476) and from a control (cv. Jack) were ground in liquid nitrogen and resuspended in 15 mL of homogenization buffer [0.25 *M* sucrose, 0.2 *M* NaCl, 0.1 *M* Tris-HCL (pH 9.0), 10% (w/v) sarkosyl, 100 m*M* β-mercaptoethanol] followed by the addition of 15 mL of phenol/chloroform (5:1, pH 4.5) (Ambion, Austin, TX) and, in turn, 1 mL of 3 *M* sodium acetate (pH 5.2). After precipitation in a final concentration of 2M LiCl, the RNA pellet was resuspended in RNase-free water. Fifteen micrograms of total RNA was run on 8% (v/v) formaldehyde gel with 1× MOPS (10× MOPS: 200 m*M* MOPS, 50 m*M* sodium acetate, 10 m*M* EDTA, pH 7) as running buffer at 115 V for 2.5 h. The gel was blotted using capillary action (Sambrook et al., 1989) and 10× SSC. Production of a P34 specific probe and hybridization were as described in Southern analysis.

Cloning and Sequencing

Primers were designed to amplify the P34 open reading frame from the known soybean P34 1.1-kb cDNA sequence (GenBank accession J05560), the same as was used in Southern probe production. PCR conditions were as described above for probe production. PCR was performed as described above in the production of a probe for hybridization analysis. Amplification products were purified with a gel extraction kit (Qiagen) and cloned into the Topo vector (Invitrogen) both according to the manufacturer's instructions. Transgenic bacterial colonies were selected on 50 μg/mL kanamycin-containing media overnight at 37°C and multiplied for plasmid extraction (Miniprep Kit, Qiagen) and analyzed for the presence of an insert by digestion with *Eco*RI. Sequencing reactions were subsequently performed on plasmids identified to contain the desired inserts. Sequencing reactions consisted of 500 ng plasmid DNA, 2 μL BigDye version 3 (ABI Foster CA), and 1 μL 100 ng/μL primer (IDT technologies, Coralville, IA). To get a good coverage of the sequence, universal primers M13 forward and reverse as well as the following internal gene-specific primers were used: forward primer bases 251 to 278 (5′-ATCACCCCATTCTC-ATCGTTTAGGATT-3′), forward primer bases 331 to 354 (5′- CTCCCAAGGATCTCTCGCAGCAA-3′), reverse primer bases 902 to 934 (5′-GCTGAACCATAACCCACAAGTAA-AACAAAGTG-3′) and reverse primer 964 to 988 (5′-TCTC-CCCAATCTTCTCCCCATGAA-3′). (Bases are numbered according to GenBank accession J05560). For each of the two *G. max* null accessions, four clones were sequenced and analyzed to compile an accurate overall sequence. Sequencing cycling parameters were: 94°C for 30 s, 50°C for 15 s, and 60°C for 1.5 min repeated 34 cycles. The resultant amplification products were cleaned by Sigmaspin post-reaction clean-up columns (Sigma) dried in a speed vacuum and analyzed (Washington University, St. Louis, MO). Sequence alignments were performed by Vector NTI program.

RESULTS

Survey of the Soybean Collection

The rapid method of dot blotting samples enabled us to screen about 100 seed samples in a day which included the protein extraction and dot blotting. From the screen, 12 accessions that were low or null for P34 were identified consisting of two *G. max*, seven *G. soja*, and three wild perennial *Glycine* species (Table 1). The P34 nulls obtained from the immuno-dot blots were subsequently analyzed by SDS-PAGE/western blot and confirmed to have lower than control levels of P34 allergen. Figure 1 shows the 12 accessions of interest in both 1-dimensional protein profile and western blot. All seven *G. soja* accessions appeared to have low levels of the P34 protein compared with cv. Kunitz (Fig. 1Aii), while the two *G. max* and the three wild perennial accessions seem to be devoid of the allergen protein (Fig. 1Cii and 1Bii, respectively).

Immunoblot Analysis of Selected Accessions

The two accessions that emerged from the screens of particular interest are the *G. max* P34 nulls: PI603570A and PI567476. To determine if a large scale variation in protein production was the underlying cause of the near-absent P34 levels in the seeds of these two accessions, 2-D SDS/PAGE and immuno-blot analysis was performed on total protein isolated from the seeds from these two accessions and compared with that of control (cv Kunitz). As can be seen in Fig. 2Aii, a standard level of the allergen P34 is readily detectable at approximately 34-kDa molecular weight and in the acidic pH range. In contrast, both *G. max* null accessions demonstrated considerably reduced levels of the P34 allergen (Fig. 2 Bii and Cii). The minuscule levels of the P34 allergen detected in the mature seeds of the two null accessions did not occur with a corresponding vast alteration in the seed proteome as can be seen in the 2-D SDS/PAGE gels stained with Coomassie blue (Fig. 2Ai for control proteome and Fig. 2Bi and 2Ci for P34 null accessions proteomes).

Table 1. The total number of *Glycine* accessions screened for P34.

	No. screened	P34 (+)	P34 (−)
Glycine max			
Plant introductions	**13199**	**13197**	**2†**
Ancestral varieties	**91**	**91**	**0**
Recent cultivars	**849**	**849**	**0**
Glycine soja	**1114**	**1107**	**7‡**
Wild perennial *Glycine* spp.	**1013**	**1010**	**3§**
Total	**16266**	**16254**	**12**

† PI 567476, PI 603570A.

‡ PI 407128, PI 407230, PI 407255, PI 407259, PI 407262, PI 407264, PI 407270.

§ PI 446991 (*G. tomentella*), PI 509462 (*G. clandestina*), IL 0956 (*G. clandestina*), (University of Illinois number).

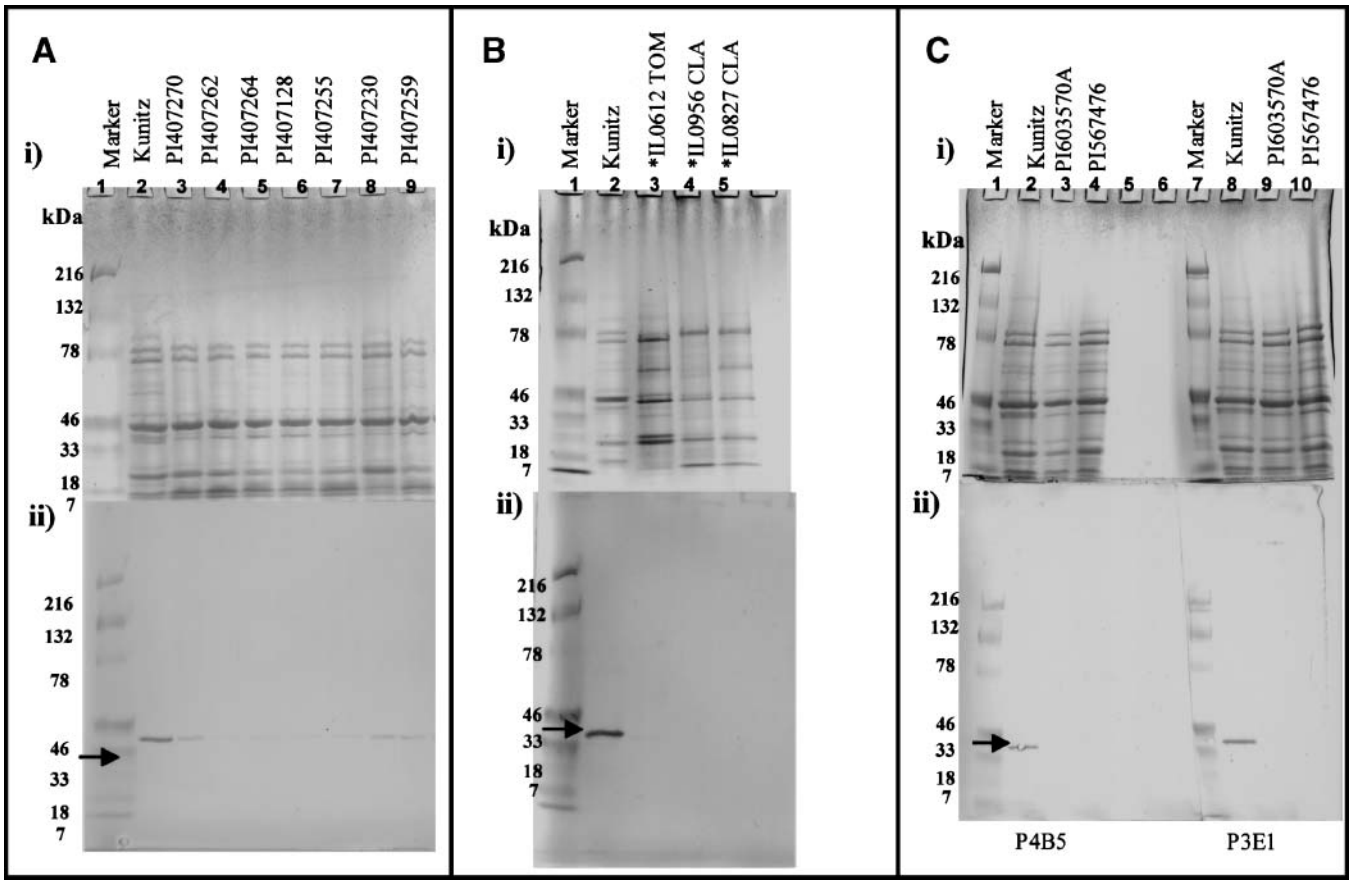

Fig. 1. One-dimensional protein profile and western blot analysis of P34-low expressing accessions. Total protein was extracted from mature seeds, separated on a 4 to 18% (w/v) gradient gel, stained with Coomassie blue [in each case upper figure (i)], and a replicate gel blotted for immunodetection using the P34 specific monoclonal antibody. Control cv. Kunitz is compared with seven accessions of *G. soja* (A) three wild perennials, * IL = University of Illinois no. IL0827 (PI509462), IL 0612 (PI 446991), IL 0956 (PI not known) (B) and two *G. max* accessions (C).

These results indicate that the reduction, to the extend of virtual elimination, of the major seed allergen P34 in the two *G. max* accessions, PI603570A and PI567476, is likely not due to a mutation in something that caused massive expression differences in the seed protein profile, such as a transcription factor, but rather, that the mutation likely only effects the production of the P34 protein.

Southern Genomic Blot Analysis of the P34 Nulls

Southern blot analysis was performed to elucidate if a genome level mutation, a large insertion or deletion perhaps, in the P34 gene was the underlying cause of the near-absent P34 protein levels in the seeds of the two *G. max* accessions of interest. From analysis of published P34 sequences (GenBank accessions AB013289 and J05560) it was predicted that the restriction enzyme *Hae*III would cleave the open reading frame of the P34 gene once at position 5005 and again 5′ upstream of the gene at position 3763 (numbers according to genomic clone accession AB013289), thereby liberating a fragment of predicted 1242 bp size. As can be seen from Fig. 3, cv. Jack was used as control as is commonly used in soybean genomics and contains standard levels of the P34 allergen in its seed. When genomic DNA was digested with *Hae*III and probed with 1.1-kb open reading frame of the P34 gene, the expected approximately 1.2-kb band is seen in cv. Jack (Fig. 3, Lane 1), but the two *G. max* P34 null accessions show a slightly larger than 1.2-kb band (Fig. 3, Lane 2 and 3). Also, the other band seen in the *Hae*III digest is approximately 1.6 kb in cv. Jack (Fig. 3, Lane 1) but larger than that in the two *G. max* accessions studied (Fig. 3; Lanes 2 and 3). Likewise the digest of *Eco*RI liberated an unexpected small fragment, approximately 650 bp, as the *Eco*RI restriction enzyme is not known to cut within the open reading frame of the P34 gene. The appearance of this band could be due to the infamous star activity of *Eco*RI, and hence the expected band size from this digest cannot be determined. As with the findings from the *Hae*III digest, digested genomic DNA sample of cv. Jack has a slightly smaller molecular weight than does either band liberated in the two *G. max* accessions when subjected to the star activity of restriction enzyme *Eco*RI. The Southern blot analysis indicates that there is some small (less than 300 bp) alteration either in the open reading frame of the P34 gene or in its somewhat immediate surroundings.

Northern RNA Blot of the P34 Null

Because the Southern blot indicated that there might be a small alteration in the DNA in or around the P34 gene, northern blot analysis of total RNA isolated from

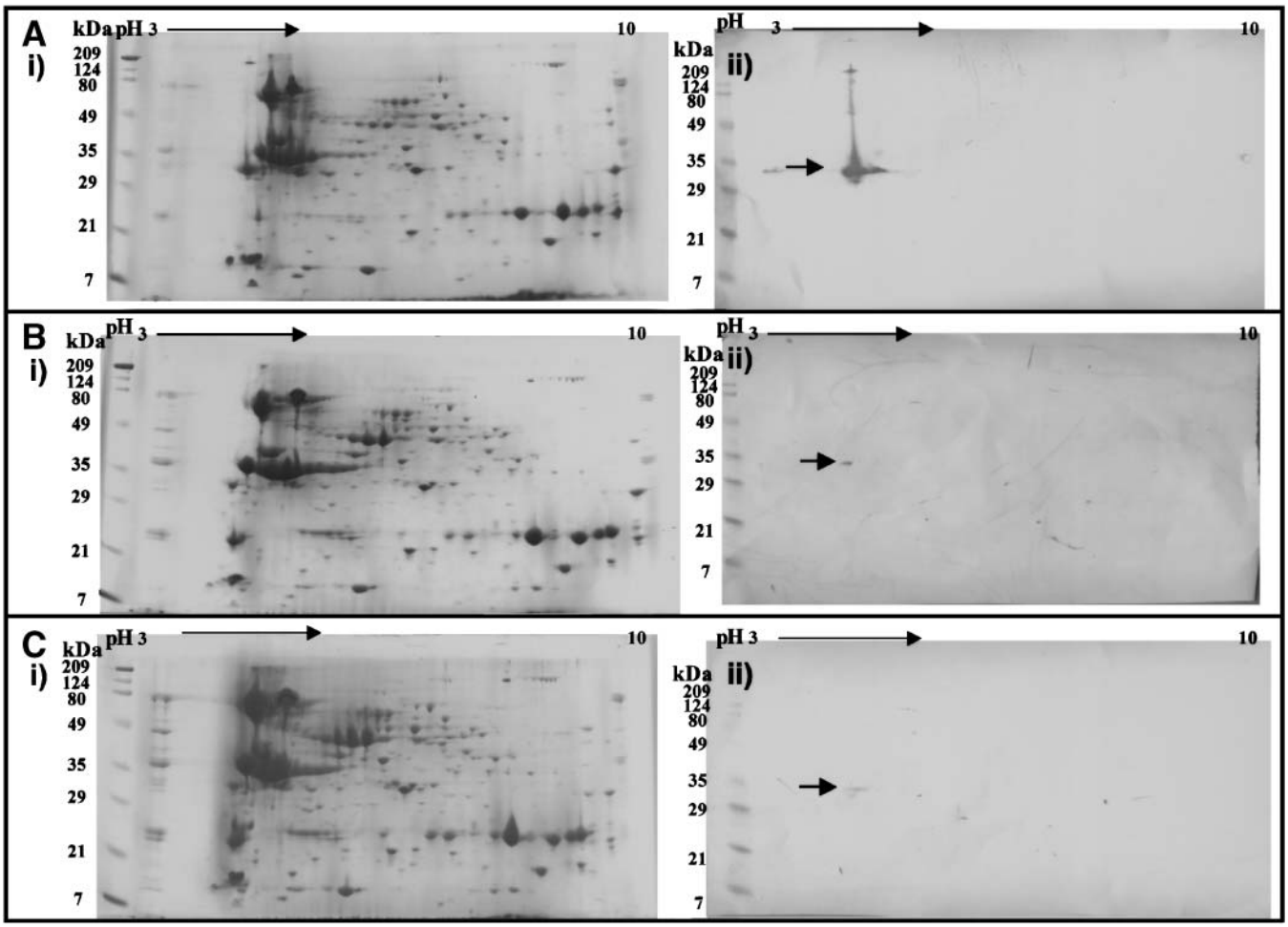

Fig. 2. Proteome analysis of two *G. max* P34-null accessions by 2-dimensional protein analysis. Proteins were separated using a pH gradient from 3.0 to 10.0 in the first dimension and 4 to 16% (w/v) gels in the second. Total proteins were detected by Coomassie blue stain [in each case, upper gel (i)] and a western blot performed on a replicate gel using P34-specific monoclonal antibody [in each case lower blot (ii)]. The cv. Kunitz (A) was used as a positive control and compared with the two P34-null accessions *G. max* PI603570A (B) and *G. max* PI567476 (C).

green mature cotyledons was performed to see if the production of the P34 transcript was altered in the two *G. max* accessions of interest. As seen in Fig. 4, both the level of expression and length of the P34 transcript in the two *G. max* accessions studied (Lane 2 and 3) are comparable to that of standard levels of cv. Jack (Lane 1). This indicates that transcription of the P34 open reading frame is complete and to the same extent in the two null accessions as in control soybean seed.

Sequence Analysis of the Genes of the *G. max* P34 Nulls

The P34 open reading frame from the two *G. max* accessions of interest were cloned and compared with that of the known published full-length P34 cDNA sequence. Figure 5 depicts a predicted amino acid alignment of the two P34 null *G. max* accessions identified in this study to that of the known amino acid sequence of the open reading frame of P34 (GenBank accession J05560). In the 1.14-kb open reading frame, both accessions were found to have six single nucleotide alteration in the sequence (GenBank accession DQ324851), which are predicted to result in four amino acid substitutions. Specifically, at position 218, a C-T transition resulted in serine[72] to be converted into a leucine; position 477 changed from G-C transversion, resulting in arginine[158] to be converted into serine; position 593 a C-G transversion, resulting in serine[197] to be converted into cysteine; and at position 606 a G-C transversion, resulting in glutamine[202] to be converted into histidine. There are also two silent mutations at nucleotides 645 and 825 which are not predicted to result in an amino acid change and hence of no interest (numbering as GenBank J05560).

DISCUSSION

Food allergies are rapidly growing in the industrialized countries with a doubling rate of less than 10 yr (Gupta et al., 2003). Reason for the rapid rise in food sensitivity in the population of industrialized but not developing countries is a subject of debate and speculation. Soybean is among the eight most significant food allergens and with its increasing use in processed and prepared foods integral to the diets in the industrialized world, the growing soybean sensitivity in both infants and adults results in a broadly significant dietary impact.

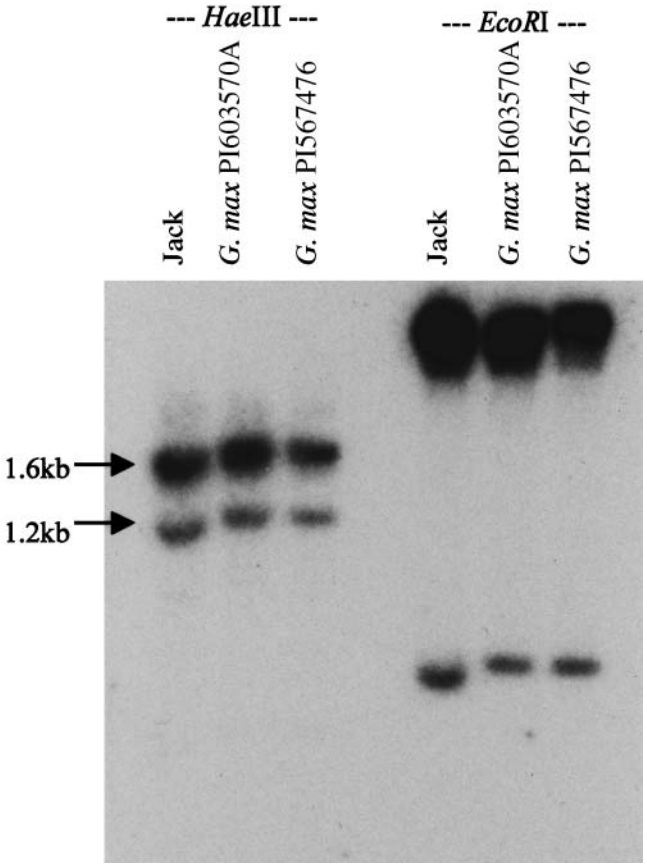

Fig. 3. Southern blot analysis of total DNA from *G. max* P34 null accessions. Samples were digested with either *Hae*III or *Eco*RI and probed with a ^{32}P-labeled 1.1-kb P34 amplicon. Banding pattern was compared with that of a control cv. Jack.

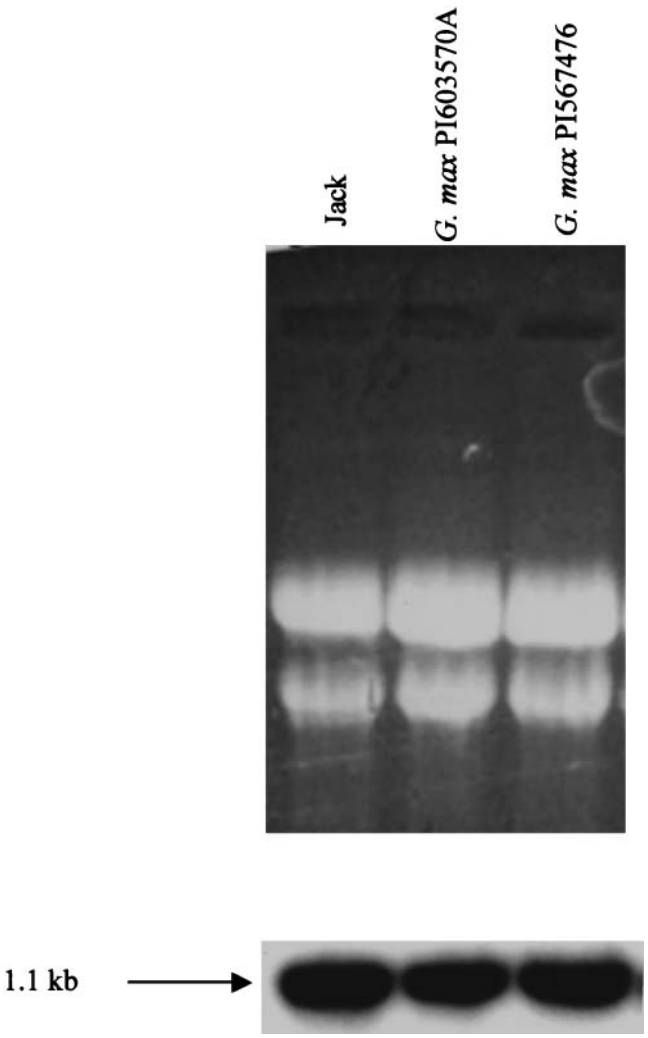

Fig. 4. Northern blot analysis of *G. max* P34 null accessions. Total RNA was isolated from mature cotyledons and probed with a ^{32}P-labeled 1.1-kb P34 amplicon. Upper panel is the ethidium bromide stained gel for a loading control. The cv. Jack was used as a control.

There are three general approaches to manage food allergies: avoidance, treatment, and low allergen-content food. Currently, the primary approach in managing food allergies is avoidance and with increased labeling requirements recently enacted into U.S. law, avoidance will be more easily practiced. However, avoidance is still very difficult when the consumer has no control over the ingredients such as in restaurants where the consumer has no access to the supplier's labels. With thousands of processed foods containing soybean in the U.S. and European market, avoidance greatly reduces the food options for sensitive individuals. There are a number of strategies for medical treatment of food allergies being developed including new drugs and immunotherapy. These are showing promise in tests; however, when or if these will be available remains uncertain. The third general approach is to produce and/or discover soybeans and other allergenic foods that have reduced, or are null for allergen content. Such varieties of an allergen null–reduced soybean could be especially valuable for uses in baby formula where the soy-content constitutes 100% of the diet and where there are limited alternatives if the infant is also milk protein sensitive.

This project is a part of an experimental approach to mitigate food allergies by identifying low allergen lines of soybean. Herman et al. (2003) previously showed that biotechnology is technically suitable to greatly reduce endogenous allergenicity. However, regulatory requirements and other factors make it difficult to introduce genetically modified allergen-reduced foods into the market, especially for uses in infant formula or food where both producers and consumers have resisted using any biotechnology-derived components. Elimination of the P34 protein would reduce the majority of the soybean allergen content and the genetic modification experiments indicate that its elimination of P34 is a tolerated change. Two previous soybean surveys of limited lines failed to identify any P34 null or reduced accessions (Ogawa et al., 2000; Yaklich et al., 1999).

After surveying the entire USDA soybean collection, we identified only two nulls in *G. max* for the soybean seed allergen P34. The two *G. max* accessions we have identified contain only a very small amount of P34 protein shown in the 2D IEF/SDS-PAGE immunoblots. The soybean is an archeopolyploid and contains additional genes derived from an unknown genome. These additional genes can be weakly expressed and we speculate that the low level of P34 protein detected is located in

```
G.max PI567476      MGFLVLLLFS LLGLSSSSSI STHRSILDLD LTKFTTQKQV SSLFQLWKSE
G.max PI603570A     MGFLVLLLFS LLGLSSSSSI STHRSILDLD LTKFTTQKQV SSLFQLWKSE
G.max P34 cDNA      MGFLVLLLFS LLGLSSSSSI STHRSILDLD LTKFTTQKQV SSLFQLWKSE

                    51                                                  100
G.max PI567476      HGRVYHNHEE EAKRLEIFKN NLNYIRDMNA NRKSPHSHRL GLNKFADITP
G.max PI603570A     HGRVYHNHEE EAKRLEIFKN NLNYIRDMNA NRKSPHSHRL GLNKFADITP
G.max P34 cDNA      HGRVYHNHEE EAKRLEIFKN NSNYIRDMNA NRKSPHSHRL GLNKFADITP
                                          *
                   150
G.max PI567476      QEFSKKYLQA PKDVSQQIKM ANKKMKKEQY SCDHPPASWD WRKKGVITQV
G.max PI603570A     QEFSKKYLQA PKDVSQQIKM ANKKMKKEQY SCDHPPASWD WRKKGVITQV
G.max P34 cDNA      QEFSKKYLQA PKDVSQQIKM ANKKMKKEQY SCDHPPASWD WRKKGVITQV

                    151                                                 200
G.max PI567476      KYQGGCGSGW AFSATGAIEA AHAIATGDLV SLSEQELVDC VEESEGCYNG
G.max PI603570A     KYQGGCGSGW AFSATGAIEA AHAIATGDLV SLSEQELVDC VEESEGCYNG
G.max P34 cDNA      KYQGGCGRGW AFSATGAIEA AHAIATGDLV SLSEQELVDC VEESEGSYNG
                          *                                              *
                    201                                                 250
G.max PI567476      WHYQSFEWVL EHGGIATDDD YPYRAKEGRC KANKIQDKVT IDGYETLIMS
G.max PI603570A     WHYQSFEWVL EHGGIATDDD YPYRAKEGRC KANKIQDKVT IDGYETLIMS
G.max P34 cDNA      WQYQSFEWVL EHGGIATDDD YPYRAKEGRC KANKIQDKVT IDGYETLIMS
                     *
                    251                                                 300
G.max PI567476      DESTESETEQ AFLSAILEQP ISVSIDAKDF HLYTGGIYDG ENCTSPYGIN
G.max PI603570A     DESTESETEQ AFLSAILEQP ISVSIDAKDF HLYTGGIYDG ENCTSPYGIN
G.max P34 cDNA      DESTESETEQ AFLSAILEQP ISVSIDAKDF HLYTGGIYDG ENCTSPYGIN

                    301                                                 350
G.max PI567476      HFVLLVGYGS ADGVDYWIAK NSWGEDWGED GYIWIQRNTG NLLGVCGMNY
G.max PI603570A     HFVLLVGYGS ADGVDYWIAK NSWGEDWGED GYIWIQRNTG NLLGVCGMNY
G.max P34 cDNA      HFVLLVGYGS ADGVDYWIAK NSWGEDWGED GYIWIQRNTG NLLGVCGMNY

                    351
G.max PI567476      FASYPTKEES ETLVSARVKG HRRVDHSPL.
G.max PI603570A     FASYPTKEES ETLVSARVKG HRRVDHSPL.
G.max P34 cDNA      FASYPTKEES ETLVSARVKG HRRVDHSPL.
```

Fig. 5. Comparison of predicted amino acid sequences of the seed allergen P34 from two *G. max* P34 null accesssions. The pre-determined pre-pro peptide sequence is underlined. Translation of the P34 cDNA obtained from GenBank accession J05560. * denotes an alteration in the amino acid sequence.

the other genome that is greatly suppressed. Even considering the small level of protein that may be derived from the other genes, the P34 content of these two lines is less than 1% of the standard level that is consistently observed across the soybean collection (Yaklich et al., 1999).

Both the *G. max* null accessions express an apparently full length mRNA of the same size as that of soybean lines that do accumulate P34. Similarly, Southern blots of genomic DNA probed with a P34 indicates the genomic fragments containing P34 is the same size in both the controls containing P34 and the two null accessions. This suggests that the defect of the *G. max* null accessions is one of translation, since the gene is transcribed producing a full length mRNA. To elucidate the cause for the lack of P34 accumulation in the nulls, the open reading frame was cloned and sequenced from cDNA produced from total RNA isolated from mature cotyledons of the two P34 null accessions and compared with previously identified P34 sequence. The data show that the two null accessions have identical predicted amino acid sequences, indicating that they are most likely derived from the same mutant. The predicted amino acid sequence of the P34 null accessions differ in four residues from the control sequence. Albeit, one of these alterations, serine[72] to leucine, occurs in the pre-propolypeptide portion of the molecule (Kalinski et al., 1990, 1992), and so it likely is not involved in the mechanism by which the mature P34 protein is not functional in the seeds of the two isolated null accessions. Since two independent immuno-detection screens were performed to identify the two P34 null *G. max* accessions on the basis of two different monoclonal antibodies, although the exact epitopes recognized by the monoclonal antibodies are not known, it has been previously determined that they are at least 3 kDa apart (Herman et al., 1990). Hence the three amino acid substitutions found in the mature P34 protein in the two null *G. max* accessions are likely too close to one another in the sequence to be altering a single epitope region. Therefore, the most likely scenario for the lack of the P34 allergen in the two P34 null accessions is the notable

change from serine[197] to cysteine. Since cysteine residues are known to be of paramount importance in disulfide bond formation, which in turn ensures correct overall three-dimensional structure, it is postulated that with the addition of a cysteine residue the protein misfolds and hence would be inactive in the soybean seed. Numerous studies have been performed to show the importance of cysteine residues in protein function by way of correct disulfide bond formation (Garofalo et al., 1991; Tarnow et al., 2003).

Future work includes screening of the P34 null lines to test for allergenic reaction using sera of soy-sensitive patients and using animal models. The accessions we have identified have potential to overcome some of the impediments to using low-allergen soybeans in processed foods. The prospect for biotechnology to introduce food allergens is widely used by critics of genetic modification of foods as one of their primary arguments against this technology. By incorporating these newly identified non-genetically modified nulls for the major soybean allergen into elite germplasm via conventional breeding, it bypasses the genetic modification problem and presents an opportunity to develop a desirable commodity. The results of this research are crucial in areas where genetically modified soybeans cannot be used without controversy, such as in baby food and formula. Soybean lacking P34 has an important role because of its implications as a paradigm to use biotechnology as a means to eliminate allergens from food that benefits humans and in animal feed. Also, because these soybeans are naturally occurring accessions obtained from the soybean collection the P34 null trait can be crossed into other germplasm to produce elite or stacked trait lines without any biotechnology-derived component.

The P34 null trait from the *G. max*, as well as the products of introgression, could be stacked with other desirable traits in improving soybean quality. This research is the first broad, systematic, comprehensive analysis of a major allergen of any crop that has been undertaken with the goal of introgressing the allergen-null trait into elite breeding lines. In addition, this research project demonstrates the value of maintaining extensive collections of major crop domesticates and their wild relatives to provide additional value-added traits that can be incorporated in products needed by consumers.

ACKNOWLEDGMENTS

This research was supported in part by the Illinois Agricultural Experiment Station and by an IMBA grant (Illinois-Missouri Biotechnology Alliance) awarded jointly to TH and EH). We wish to thank Ms. Jo Weatherspoon (USDA/ARS, Urbana) and Ms. Nhung Pham (University of Illinois hourly technical assistant) who packaged by hand about 16,266 accessions of *Glycine* for P34 analysis. We thank Dr. Randall Nelson USDA/ARS curator for his cooperation.

REFERENCES

Bradford, M.M. 1976. A rapid and sensitive method for the quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72:248–254.

Burks, A.W., H.B. Castell, S.C. Fiedorek, L.W. Williams, C. Connaughton, and J.R. Brooks. 1989. Enzyme-linked immunosorbant assay and immunoblotting determination of antibody response to major component proteins of soybeans in patients with soy protein intolerance. J. Pediatr. Gastroenterol. Nutr. 8:195–203.

Cantani, A., and P. Lucenti. 1997. Natural history of soy allergy and/or intolerance in children, and clinical use of soy-protein formulas. Pediatr. Allergy Immunol. 8:59–74.

Cashmore, A.R., M.K. Broadhurst, and R.E. Gray. 1978. Cell-free synthesis of leaf protein: Identification of an apparent precursor of the small subunit of ribulose-1,5-bisphosphate carboxylase. Proc. Natl. Acad. Sci. USA 75:655–659.

Cheng Ji, C. Boyd, D. Slaymaker, Y. Okinaka, E.M. Herman, and N.T. Keen. 1998. Purification and characterization of a 34 kDa syringolide binding protein from soybean. Proc. Natl. Acad. Sci. USA. 95:3306–3311.

Church, G.M., and W. Gilbert. 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA 81:1991–1995.

Doyle, J.J., and J.L. Doyle. 1990. Isolation of DNA from small amounts of plant tissue. BRL Focus 12:13–15.

Foucard, T., and Y.I. Malmheded. 1999. A study on severe food reactions in Sweden- is soy protein an underestimated cause of food anaphylaxis? Allergy 54:261–265.

Garofalo, S., E. Vuorio, M. Metsaranta, R. Rosati, D. Toman, J. Vaughan, G. Lozano, R. Mayne, J. Ellard, W. Horton, and B. Crombrugghe. 1991. Reduced amounts of cartilage collagen fibrils and growth plate anomalies in transgenic mice harboring a glycine-to-cysteine mutation in the mouse type II procollagen alpha 1-chain gene. Proc. Natl.Acad. Sci. (USA). 88: 21:9648–9652

Gupta, R., A. Sheikh, D. Strachan, and H.R. Anderson. 2003. Increasing hospital admissions for systemic allergic disorders in England: Analysis of national admissions data. BMJ 327:1142–1143.

Hajduch, M., A. Ganapathy, J.W. Stein, and J.J. Thelen. 2005. A Systematic proteomic study of seed filling in soybean. Establishment of high-resolution two-dimensional reference maps, expression profiles and an interactive proteome database. Plant Physiol. 137:1397–1419.

Helm, R.M., G. Cockrell, E.M. Herman, A.W. Burks, H.A. Sampson, and G.A. Bannon. 1998. Cellular and molecular characterization of a major soybean allergen. Int. Arch. Allergy Immunol. 117:29–37.

Helm, R.M., G. Cockrell, C.M. West, E.M. Herman, H.A. Sampson, G.A. Bannon, and A.W. Burks. 2000. Mutational analysis of the IgE-binding epitopes of P34/Gly m1. J. Allergy Clin. Immunol. 105: 378–384.

Heppel, L.M., J.W. Sissons, and H.E. Pedersen. 1987. A comparison of the antigenicity of soya-bean-based infant formulas. Br. J. Nutr. 58: 393–403.

Herian, A.M., S.L. Taylor, and R.K. Bush. 1993. Allergenicity of various soybean products as determined by RAST inhibition. J. Food Sci. 58:385.

Herman, E.M., D.L. Melroy, and T.J. Buckhout. 1990. Apparent Processing of a Soybean Oil Body Protein Accompanies the Onset of Oil Mobilization. Plant Physiol. 94:341–349.

Herman, E.M. 2004. Allergenic responses to legume proteins. p. 326–347. *In* C. Brummer and R. Wilson (ed.) Genomics for legume crops. AOCS Press, Champaign, IL.

Herman, E.M. 2005. Soybean allergenicity and suppression of the immunodominant allergen. Crop Sci. 45:462–467.

Herman, E.M., R. Helm, R. Jung, and A. Kinney. 2003. Genetic modification removes an immunodominant soybean allergen. Plant Physiol. 132:36–43.

Hurkman, W.J., and C.K. Tanaka. 1986. Solubilization of Plant Membrane Proteins for Analysis by Two-Dimensional Gel Electrophoresis. Plant Physiol. 81:802–806.

Kalinski, A.J., J.M. Weisemann, B.F. Matthews, and E.M. Herman. 1990. Molecular cloning of a protein associated with soybean seed oil bodies that are similar to thiol proteases of the papain family. J. Biol. Chem. 265:13843–13848.

Kalinski, A.J., D.L. Melroy, R.S. Dwivedi, and E.M. Herman. 1992. A soybean vacuolar protein (P34) related to thiol protease which is synthesized as a glycoprotein precursor during seed maturation. J. Biol. Chem. 267:12068–12076.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685.

Ogawa, T., N. Bando, H. Tsuji, H. Okajima, K. Nishikawa, and K. Sasaoka. 1991. Investigation of the IgE-binding proteins in soybeans by immunoblotting with the sera of the soybean-sensitive patients with atopic dermatitis. J. Nutr. Sci. Vitaminol. (Tokyo) 37:555–565.

Ogawa, T., H. Tsuji, N. Bando, K. Kitamura, Y.L. Zhu, H. Hirano, and K. Nishikawa. 1993. Identification of the soybean allergenic protein, Gly m Bd 30K, with the soybean seed 34-kDa oil-body-associated protein. Biosci. Biotechnol. Biochem. 57:1030–1033.

Ogawa, T., M. Samoto, and K. Takahashi. 2000. Development of hypoallergenic soybean products-strategy for the reduction of allergencity of soybean. p. 165–168. *In* Proceedings of The Third International Soybean Processing and Utilization Conference, Tsukuba, Ibaraki, Japan. 15–20 Oct. 2000. Congress Corporation, Tokyo.

Tarnow, P., T. Schoneberg, H. Krude, A. Gruters, and H. Biebermann. 2003. Mutationally induced disulfide bond formation within the third extracellular loop causes meanocortin 4 receptor inactivation in patients with obesity. J. Biol. Chem. 278:48666–48673.

Tsuji, H., N. Okada, R. Yamanishi, N. Bando, M. Kimoto, and T. Ogawa. 1995. Measurement of Gly m Bd 30K, a major soybean allergen, in soybean products by a sandwich enzyme-linked immunosorbent assay. Biosci. Biotechnol. Biochem. 59:150–151.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular cloning. A laboratory manual. Cold Spring Harbor, NY.

Yaklich, R., R. Helm, and E.M. Herman. 1999. Analysis of the distribution of the major soybean allergen in a core collection of *Glycine max* accessions. Crop Sci. 39:1444–1447.